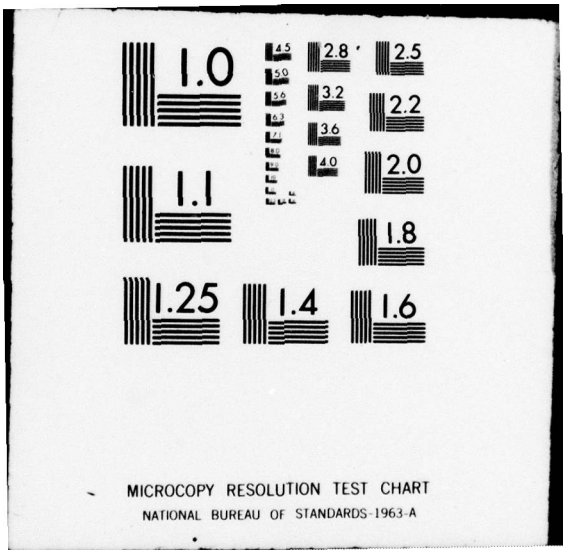
AD-A067 773 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1
TIME SERIES MODELING, SPECTRAL ANALYSIS, AND FORECASTING.(U)
MAR 79 E PARZEN N00014-78-C-0599
UNCLASSIFIED TR-N-7 NL

END
DATE FILMED
6-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843

INSTITUTE OF STATISTICS
Phone 713 - 845-3141

AD A067773

# TIME SERIES MODELING, SPECTRAL ANALYSIS, AND FORECASTING

by Emanuel Parzen

Institute of Statistics, Texas A&M University

Technical Report No. N-7

March 1979

Texas A & M Research Foundation
Project No. 3838

"Multiple Time Series Modeling and Time Series Theoretic Statistical Methods"

Sponsored by the Office of Naval Research

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited.

DDC FILE COPY

DDC
RECEIVED
APR 24 1979
C

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

## REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report No. N-7
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Time Series Modeling, Spectral Analysis, and Forecasting
5. TYPE OF REPORT & PERIOD COVERED: Technical
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Emanuel Parzen
8. CONTRACT OR GRANT NUMBER(s): N00014-78-C-0599
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Texas A&M University, Institute of Statistics, College Station, TX 77843
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Code 436, Arlington, VA 22217
12. REPORT DATE: March 1979
13. NUMBER OF PAGES: 50
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office): 26p.
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report): NA
18. SUPPLEMENTARY NOTES: NA
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Time series time domain models, Time series decomposition, Spectral analysis, Autoregressive schemes, Forecasting, Time Series Diagnostic Tests, Nonstationary time series
20. ABSTRACT (Continue on reverse side if necessary and identify by block number): A strategy for building models for an observed time series is presented in this paper. We seek to fit time domain models which can be interpreted in terms of trend and seasonal components, provide forecasts, and provide spectral estimators. Our time series modeling strategy attempts to achieve these aims by using the concepts of predictability and fitting the spectral distribution function. The approach described could be called: the autoregressive spectral method for time domain model identification of non-stationary time series (which could be abbreviated AR-SPECTRAL-TIME-ID).

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

347380

79 04 23 109

# TIME SERIES MODELING, SPECTRAL ANALYSIS, AND FORECASTING

by

Emanuel Parzen

Texas A&M University

## Abstract

A strategy for building models for an observed time series is presented in this paper. We seek to fit time domain models which can be interpreted in terms of trend and seasonal components, provide forecasts, and provide spectral estimators. Our time series modeling strategy attempts to achieve these aims by using the concepts of predictability and fitting the spectral distribution function. The approach described could be called: "the auto-regressive spectral method for time domain model identification of non-stationary time series" (which could be abbreviated AR-SPECTRAL-TIME-ID).

The paper contains the following sections: 1. Introduction; 2. The time correlation problem, time series decomposition, and whitening filters; 3. Covariance stationary time series; 4. The stages of time series modeling; 5. Stationary analysis of a time series; 6. Non-stationary analysis; 7. International airline data; 8. Metals series of Makridakis; 9. How well can stationary time series models predict?; 10. Diagnostic checks for time series models; 11. ARMA models.

Research supported in part by the Office of Naval Research (Contract N00014-78-C-0599).



## 1. Introduction

A strategy for building models for an observed time series is presented in this paper. We seek to fit time domain models which can be interpreted in terms of trend and seasonal components, provide forecasts, and provide spectral estimators. Our time series modeling strategy attempts to achieve these aims by using the concepts of predictability and fitting the spectral distribution function. The approach described could be called: "the auto-regressive spectral method for time domain model identification of non-stationary time series." It represents the approach to empirical time series analysis which I have been developing since Parzen (1967). It could be called the AR-SPECTRAL-TIME-ID method.

To understand the notion of predictability, consider a time series $Y(t)$ obeying the simplest model. It is a sum of a constant mean $c$ and white noise $\varepsilon(t)$ with variance $\sigma^2$,

$$Y(t) = c + \varepsilon(t) .$$

Given a sample of size $T$, estimators of $\hat{c}$ and $\hat{\sigma}^2$ of $c$ and $\sigma^2$ are respectively the sample mean and sample variance. The ratio $\hat{c}/\hat{\sigma}$ can be used to classify the time series into one of three states:

(i) white noise if $\hat{c}^2/\hat{\sigma}^2 < 4/T$; (one may act as if $c = 0$);

(ii) predictable if $\hat{c}^2/\hat{\sigma}^2 > (T/8)^{-1}$; (one may act as if $\sigma = 0$);

(iii) unpredictable if $4/T < \hat{c}^2/\hat{\sigma}^2 < (T/8) - 1$.

It should be emphasized that the exact criteria that $\hat{c}^2/\hat{\sigma}^2$ need satisfy to classify the time series into one of the above three states is somewhat arbitrary. The goal of this paper is to show that the first step in modeling an observed time series is to classify it into one of the above three

states (however, using suitable criteria which are in general different from those stated above for a constant in white noise).

To classify a time series into one of the states (predictable, unpredictable, white noise) one can use the same mathematical formulas and computing algorithms used to estimate the parameters of an autoregressive scheme for a stationary time series. The output of such an analysis is a best fitting order $\hat{m}$, a residual variance $\hat{\sigma}^2_{\hat{m}}$, and estimated autoregressive coefficients $\hat{a}_1, \ldots, \hat{a}_m$. We recommend that the time series be classified:

(i) white noise if $\hat{m} = 0$, or $\hat{\sigma}^2_{\hat{m}} \doteq 1$,

(ii) predictable if $\hat{\sigma}^2_{\hat{m}} \doteq 0$,

(iii) unpredictable otherwise.

An unpredictable time series is regarded as a stationary time series approximately satisfying an autoregressive scheme of order $\hat{m}$, with coefficients $\hat{a}_1, \ldots, \hat{a}_m$, and with spectral density estimated by the spectral density of the autoregressive scheme. The crucial problem of choosing the order $\hat{m}$ is solved using order determining criteria and comparisons of spectral distribution functions.

A predictable time series is regarded as being predictable because of the presence of trend and seasonal components (and therefore non-stationary). Consequently, these intuitive notions have to be given as precise definitions as possible. The operations on a time series to be used to transform a predictable (non-stationary) time series to a non-predictable (stationary) time series are those which are traditionally regarded as detrending and deseasonalizing operations.

The question will naturally be raised how the time series modeling strategy proposed in this paper compares with that developed by Box and



Jenkins (1970). The latter consists of iterations of a four-step cycle consisting of:

(0) model specification,
(1) model identification,
(2) model estimation,
(3) diagnostic checks on model adequacy.

Model specification for a time series Y(t), which could be the originally observed time series or a memoryless transformation such as the logarithm or square root, is always assumed in the Box-Jenkins method to be an ARIMA (autoregressive integrated moving average) model

$$a(B)a_s(B^s)(1 - B)^d(1 - B)^D Y(t) = c + b(B)b_s(B^s)\varepsilon(t) ,$$

where B is the backward shift operator (denoted by L in our work), $\varepsilon(t)$ is white noise, s is the seasonal period (if one is present), c is a constant, a(B) and b(B) are polynomials in B of degrees p and q respectively, $a_s(B^s)$ and $b_s(B^s)$ are polynomials in $B^s$ of degrees P and Q respectively.

The model identification stage guesstimates p, d, q, P, D, Q based on visual examination of the autocorrelation and partial autocorrelation functions of $(I - B)^d(I - B^s)^D Y(t)$ for various choices of d and D.

The model estimation stage forms estimators $\hat{a}_i$, $\hat{b}_i$, $\hat{a}_{i,s}$, $\hat{b}_{i,s}$, $\hat{c}$ of the coefficients of the model.

The diagnostic testing stages tests for white noise the residuals (assuming c = 0)

$$\hat{\varepsilon}(t) = \{\hat{b}(B)\hat{b}_s(B^s)\}^{-1}\hat{a}(B)\hat{a}_s(B^s)(1 - B)^d(I - B^s)^D Y(t) .$$

The model identification and diagnostic checking stages seem to be very subjective so that applying the Box-Jenkins model-building strategy may lead an inexperienced time series analyst to an inadequate model

(where inadequate has several meanings, such as inadequate for interpretation, or inadequate for forecasting). One important deficiency of the Box-Jenkins approach is that it does not include estimation of the spectrum among its aims of time series analysis.

Time series analysis is of increasing importance to researchers in the physical sciences, engineering sciences, biological sciences, and medicine, as well as to those in the social sciences, economics, and management science. As an interdisciplinary field, time series analysis should attempt to achieve unified approaches which will serve simultaneously the full diversity of its clients.

## 2. The Time Correlation Problem, Time Series Decomposition, and Whitening Filters

A traditional approach in time series analysis has been to regard a typical time series as composed of three parts: (a) a trend, or long term movement, (b) a seasonal variation, or oscillations about the trend of greater or less regular periodicity, and (c) a residual, or random, irregular, or unsystematic component; a primary purpose of time series analysis is then to decompose the time series into a sum of three series representing trend, seasonal, and residual.

This traditional decomposition was proposed by the earliest researchers in time series analysis; their point of view is beautifully expressed by Yule (1921):

> I hope that I have not been unduly labouring the obvious, but wish to emphasize that it is not my view alone but the view of most writers on the subject up to 1914, that the essential difficulty of the time-correlation problem is the difficulty



> of isolating for study different components in the total movement of each variable: the slow secular movement, probably non-periodic in character or, if periodic, with a very long period; the oscillations of some ten years' duration, more or less, corresponding to the wave in trade; the rapid movements from year to year which give an appearance of irregularity to the curve in a statistical chart and which may in fact be irregular or may possess a quasi-periodicity of some two years' duration; the seasonal movement within the year, and so on. It is unfortunate that the word "periodic" implies rather too much as to the character of such more rapid movements; few of us, I suppose, now believe that they are strictly periodic in the proper sense of the term, and hence the occurrence in writings on the subject of such terms as "quasi-periodic" and "psuedo-periodic." They are wave-like movements, movements which can be readily represented with a fair degree of accuracy over a moderate number of years by a series of harmonic terms but which cannot be represented in the same way, for example, by a polynomial; movements in which the length of time from crest to crest of successive waves is not constant, and in which, it may be added, the amplitude is not constant either, but would probably, if we could continue our observations over a sufficient number of waves, exhibit a frequency distribution with a fairly definite mode; to avoid the suggestion of strict periodicity and the use of the term <u>period</u> I propose to speak of them as <u>oscillations</u> of a given <u>duration</u>, the word <u>duration</u> to imply, not a fixed and constant duration, but an average only. In these terms, the problem of time-correlation may be said to be the isolation, for separate study, of oscillations of differing durations. Most writers up to 1914 - indeed all writers so far as I am aware - seem to be agreed on this.

It is still an unsolved problem how to carry out in practice for an arbitrary observed time series a decomposition into trend, seasonal or cyclic, and residual components. But a criterion that any general time series modeling strategy must fulfill is that its conceptual framework should provide an explanation or role for the continuing quest for a time series decomposition. As one reads the above quotation from Yule (1921), one's understanding of what he means could be obtained by thinking in the spectral domain. <u>Thus it seems critical that a successful approach to time series modeling employ simultaneously both the spectral domain and the time domain</u>.

A time-domain approach to the problem of mathematical definitions of trend and seasonal is provided by definitions such as the following.

By trend is meant that part of a time series which grows by rules such as its first difference is constant, or some higher difference is constant; then trend might be represented by a polynomial P(t) satisfying

$$\nabla P(t) := P(t+1) - P(t) = c, \text{ or } \nabla^n P(t) = c .$$

By seasonal with period $\lambda$ is meant that part of a time series which exactly repeats at intervals of time $\lambda$ apart; it might be represented as a function S(t) satisfying

$$\nabla_\lambda S(t) := S(t+\lambda) - S(t) = 0 .$$

Spectral analysis enters time series analysis because a function with period $\lambda$ can be represented as a linear combination of harmonic functions

$$1 , \quad \cos(n \tfrac{2\pi}{\lambda} t) , \quad \sin(n \tfrac{2\pi}{\lambda} t) , \quad n = 1, 2, \ldots .$$

In the important special case of an economic time series of monthly values, to account for seasonal variation in the time series (which by definition are phenomena with period 12 months), a seasonal S(t) has the representation

$$S(t) = A_0 + \sum_{j=1}^{5} (A_j \cos(j \tfrac{2\pi}{12} t) + B_j \sin(j \tfrac{2\pi}{12} t)) + A_6 \cos \pi t .$$

The representation of a time series as a sum of trend P(t), seasonal S(t), and irregular N(t) could be written in symbols

$$Y(t) = P(t) + S(t) + N(t) .$$



To begin with we assume that N(t) is white noise (a sequence of independent identically normally distributed random variables). To understand this decomposition, I propose that one regard P, S, and N as constructed in terms of suitable operations on the observed series Y which we denote as follows:

$$P(t) = \underline{P}Y(t) , \quad S(t) = \underline{S}Y(t) , \quad N(t) = (\underline{I} - \underline{A})Y(t),$$

where $\underline{I}$ is the identity operator. Then the representation of Y is a relation between operators:

$$\underline{I} = \underline{P} + \underline{S} + (\underline{I} - \underline{A}).$$

Next assume that $\underline{P}$ and $\underline{S}$ are constructed to be orthogonal which we write symbolically

$$\underline{P}\underline{S} = \underline{0} ,$$

where $\underline{0}$ is the zero operator. Then the representation of Y is a representation

$$\underline{I} - \underline{A} = (\underline{I} - \underline{P})(\underline{I} - \underline{S}) .$$

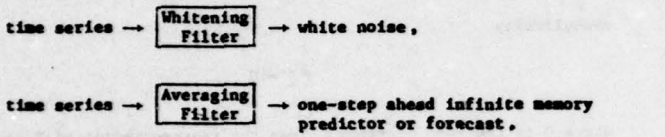
We call $\underline{I} - \underline{A}$ the <u>whitening</u> operator since it transforms the time series Y(t) to a white noise time series N(t).

We call $\underline{A}$ the averaging operator. If $\underline{A}Y(t)$ involves only past values $Y(t-1), Y(t-2), \ldots$, then $\underline{A}Y(t)$ is the best predictor of Y(t) given past values of the time series, and $\underline{A}$ is the one-step ahead infinite memory prediction operator.

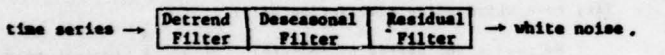
We call $\underline{I} - \underline{P}$ a detrending operator, and $\underline{I} - \underline{S}$ a deseasonal-operator.

The general problem of time series modeling (called by Yule (1921) the time-correlation problem) is first to determine a whitening filter $\underline{I} - \underline{A}$, rather than first to determine time series decomposition filters $\underline{P}$ and $\underline{S}$; the latter can be obtained as interpretations of a whitening filter. Further the forecasting problem (which can often be given solutions in terms of the filter $\underline{A}$) can be regarded as determining a whitening filter $\underline{I} - \underline{A}$. We speak of *a* whitening filter rather than *the* whitening filter because the filters fitted in practice are approximations.

A filter is represented in a diagram as a "black box" or system relating an input function to an output function:




The aim of time series modeling is not only to obtain the whitening filter but to interpret it as several filters operating successively (the output of one filter is the input of the next):




The residual filter will be described in the sequel; it transforms the residuals (which in general are stationary time series) to white noise.



## 3. Covariance Stationary Time Series

To decompose a time series $Y(t)$, one approach is to write it as a sum of a mean value function $m(t) = E[Y(t)]$ and a residual $Z(t) = Y(t) - m(t)$. By assuming a parametric formula for $m(t)$ as a sum of a polynomial $P(t)$ of known degree and a periodic function $S(t)$ of known period, one could estimate the trend and seasonal. However, even having made (most probably unjustified) assumptions almost the mean value function $m(t)$, one still has great difficulty in efficiently estimating its parameters because one needs to make assumptions about the form of the residual series $Z(t)$. One often assumes they they are jointly normally distributed random variables with zero means. Their complete probability distribution is then determined by the covariance kernel

$$K(s, t) = E[Z(s)Z(t)] = \mathrm{Cov}[Y(s), Y(t)] .$$

From a single sample $(Y(t), t = 1, 2, \ldots, T)$ of the time series, one could estimate $K(s, t)$ only by making assumptions such as the residual series is <u>stationary</u> and <u>ergodic</u>. More precisely one assumes there is a function $R(v)$ of integers $v = 0, \pm 1, \ldots$ such that

$$K(s, t) = R(|s - t|) ,$$

We call $R(\cdot)$ the covariance function of the covariance stationary time series $Y(\cdot)$.

<u>Covariance Function</u>. A covariance stationary time series (which to begin with is assumed to be normal and zero mean) is described by its covariance function denoted by $R(\cdot)$. There is no universally accepted

notation for the argument of $R(\cdot)$. I use v as the argument of $R(\cdot)$, rather than letters such as $\tau$ or j, because v looks like $\nu$ (often used to denote integer arguments) and because I regard v as inevitably paired with another variable which I regard as naturally denoted by u since u will denote a variable on the unit interval $0 \le u \le 1$. (The letters u and v clearly pair well together.)

When $Y(t)$ has zero means, the covariance function $R(v)$ is defined by

$$R(v) = E[Y(t)Y(t + v)] .$$

The correlation function $\rho(v)$ is defined by

$$\rho(v) = \frac{R(v)}{R(0)} = \mathrm{Corr}[Y(t), Y(t + v)] .$$

Two important properties of $R(v)$, and therefore of $\rho(v)$, are:

Even function: $R(-v) = R(v)$ for all v

Positive definite function: $\sum_{i,j=1}^{n} c_i c_j R(i - j) \ge 0$

for all integers n and constants $c_1, \ldots, c_n$.

Often to discuss the mathematical theory, we make an important assumption: $\rho(v)$ is summable,

$$\sum_{v=-\infty}^{\infty} |\rho(v)| < \infty .$$

In practice we will develop methods of confirming or relaxing this assumption.

Spectral Density and Power Spectral Density. The Fourier transform of $\rho(\cdot)$,



$$f(u) = \sum_{v=-\infty}^{\infty} e^{2\pi i u v} \rho(v) , \qquad 0 \le u \le 1 ,$$

will be called here the spectral density of the time series $Y(\cdot)$. The Fourier transform of $R(\cdot)$, equal to $R(0)f(u)$, will be called the power spectral density of the time series. In the physical sciences and engineering, the argument of the power spectral density function is often denoted by "f" to denote frequency, and the function itself is denoted P or S. Time series analysts often use $\omega$ as the argument of f.

The function $f(u)$ is actually defined for all u in $-\infty < u < \infty$. However, it is periodic with period 1, and its domain can be taken to be either $-0.5 \le u \le 0.5$ or $0 \le u \le 1$. The interval $-0.5 \le u \le 0.5$ is customary in the engineering literature, but only the sub-interval $0 \le u \le 0.5$ has physical significance. The deterministic time series

$$Y(t) = \cos 2\pi u t \quad \text{or} \quad Y(t) = \sin 2\pi u t$$

have period $1/u$; thus frequency u varies from 0 to 0.5, period varies from 2 to $\infty$. We adopt the interval $0 \le u \le 1$ as the basic domain of $f(\cdot)$. We plot it on the interval $0 \le u \le 0.5$, since it is an even function.

The spectral density has two important mathematical properties

Even function: $f(u) = f(-u)$,

Nonnegative function: $f(u) \ge 0$.

Spectral Distribution Function. A correlation function $\rho(v)$ always has the representation

$$\rho(v) = \int_0^1 e^{2\pi i u v} \, dF(u) , \qquad v = 0, \pm 1, \pm 2, \ldots ,$$

where $F(u)$, $0 \leq u = 1$, is a distribution function called the spectral distribution function. Like a probability distribution function, $F(u)$ can be uniquely written as sum

$$F(u) = F_{ac}(u) + F_{d}(u) + F_{sc}(u)$$

of three bounded non-decreasing functions which are respectively: absolutely continuous, in the sense that

$$F_{ac}(u) = \int_0^u f(u')du' , \quad 0 \leq u \leq 1 ;$$

discrete or purely discontinuous, in the sense that

$$F_{d}(u) = \sum_{0 \leq u' \leq u} J(u')$$

where $J(u) = F(u + 0) - F(u - 0)$ is the jump at $u$ of $F(\cdot)$; and singular continuous (a part we always assume vanishes). Spectral distribution functions play an important role in our approach to empirical modeling of time series.

White noise (a sequence of independent identically distributed random variables) is stationary with correlation function

$$\rho(v) = 0 \quad \text{if} \quad v \neq 0 ,$$

spectral density function,

$$f(u) = 1 \quad \text{for all } u ,$$

and spectral distribution function

$$F(u) = u , \quad 0 \leq u \leq 1 .$$



## 4. The Stages of Time Series Modeling

To fit a model to a time series $Y(\cdot)$ we distinguish the following stages and recommend the following strategy:

1. Transformation without memory. If the data does not appear to be normally distributed, and in particular if its distribution appears to have long tails, one may want to perform a memoryless transformation such as logarithm or square root. Experience indicates that such transformations often do not change the overall shape of the time series model fitted. Consequently this stage is optional. However, one should always examine the probability distribution of the time series. One could use quantile-box plots and other techniques of non-parametric statistical data modeling introduced by Parzen (1979).

2. Regression residuals (transformation to stationarity). When the time series clearly has trend or seasonal components one may want to represent them as a mean value function $m(t)$ with a finite number of parameters to be estimated by regression methods. One then models the regression residual time series $Y(t) - \hat{m}(t)$. We almost always subtract out the sample mean $\bar{Y}$ before starting the stationary stages of time series analysis described in Section 5.

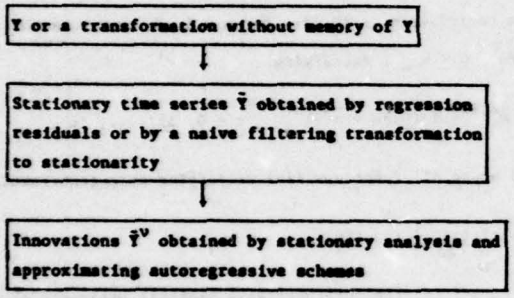
3. Naive filtering transformation to stationary. When the observed time series $Y(\cdot)$ cannot be regarded as stationary, one may be able to select a naive filtering operator (such as first difference $\nabla$ or twelfth difference $\nabla_{12}$), such that the output time series $\tilde{Y}(t)$ is stationary; in symbols,

$$Y \rightarrow \boxed{\begin{matrix}\text{Naive} \\ \text{Filter}\end{matrix}} \rightarrow \tilde{Y}$$

I strongly recommend that $\tilde{Y}$ be defined not to be as white as possible but to be as close to non-stationary as possible (while still passing for stationary).

4. Stationary Analysis, transformation to innovations. This stage determines a model for the stationary time series $\tilde{Y}$ as a whitening filter whose output is denoted $\tilde{Y}^{\nu}$, called the innovations series of $\tilde{Y}$. Parzen (1979) argues that $\tilde{Y}^{\nu} = Y^{\nu}$, the innovations series of Y.

5. Prediction and reconstruction. The general process of transforming a general time series Y to a white noise series is described symbolically as follows.




From $\tilde{Y}^{\nu}$ one can form one-step ahead infinite memory predictors of $\tilde{Y}(t)$ and Y(t) which we denote $\tilde{Y}^{\mu}(t)$ and $Y^{\mu}(t)$. A measure sometimes adopted for evaluating how successful the modeling process has been is the mean absolute percentage error

$$\text{MAPE} = \sum_{t=1}^{T} \left| \frac{Y(t) - Y^{\mu}(t)}{Y(t)} \right|$$

A relation of such measures to MSE (mean square error of one-step ahead prediction) is discussed at the end of Section 7.



## 5. Stationary Analysis of a Time Series

The so-called stationary analysis of a sample $\{Y(t), t = 1, 2, \ldots, T\}$ to be described in this section is really a generalized harmonic analysis, and does not require that the time series be "stationary" in fact. Whether the time series can be regarded as stationary will be one of the conclusions to be drawn from the analysis. A stationary analysis has as its input $\{Y(t), t = 1, \ldots, T\}$, and has as its output the following printer output.

1. Plot of the time series, mean $\bar{Y}$, variance R(0).
2. Periodogram (or raw spectral density) $\tilde{f}(u)$, $0 \leq u \leq 0.5$, defined by
$$\tilde{f}(u) = \left| \sum_{t=1}^{T} Y(t) e^{-2\pi i u t} \right|^2 + \sum_{t=1}^{T} Y^2(t)$$
and computed for u equal to a multiple of 1/Q, where Q is specified so that all frequencies corresponding to interesting frequencies appear in the printout. For monthly data with T = 144 one might choose Q = 120. The periodogram is not a consistant estimator of the true spectral density f(u) of a stationary time series. This fact is reflected in the fact that the graph of $\tilde{f}(u)$ is very wiggly.
3. Cumulative periodogram (or raw spectral distribution function)
$$\tilde{F}(u) = 2 \int_0^u \tilde{f}(u')du' , \quad 0 \leq u \leq 0.5$$
It provides a basis for comparing proposed smooth spectral density functions $\hat{f}(u)$; one desires the corresponding spectral distribution functions $\hat{F}(u)$ to "parsimoniously match" $\tilde{F}(u)$. It should be noted that we have defined $\tilde{F}(u)$ in such a way that $\tilde{F}(0.5) = 1$; this definition differs by a factor from the definition of F in section 3 which implies F(0.5) = 0.5.

4. Correlation function

$$\hat{\rho}(v) = \sum_{t=1}^{T-v} Y(t)Y(t+v) \div \sum_{t=1}^{T} Y^2(t)$$

$$= \int_0^1 e^{2\pi iuv}\, \tilde{f}(u)du$$

$$= 2\int_0^{0.5} \cos 2\pi uv\, \tilde{f}(u)du .$$

By using Fast Fourier Transform techniques one can compute $\hat{\rho}(v)$ for $v = 1, 2, \ldots, M$ where M could be any integer satisfying $M \leq T - 1$. I recommend always taking $M \geq \frac{1}{3}T$, and often one should choose $M \geq \frac{2}{3}T$.

5. Parzen window spectral densities

$$\tilde{f}_M(u) = \sum_{|v|\leq M} k(\tfrac{v}{M})\hat{\rho}(v)e^{-2\pi iuv}$$

where

$$k(t) = 1 - 6t^2 + 6|t|^3, \qquad |t| \leq 0.5$$

$$= 2(1 - |t|)^3 , \qquad 0.5 \leq |t| \leq 1$$

$$= 0 , \qquad 1 \leq |t| .$$

One computes $\tilde{f}_M(u)$ for 2 or 3 values of M, chosen approximately equal to $\frac{1}{6}T$, $\frac{1}{3}T$, and $\frac{2}{3}T$.

As a method of estimating the true spectral density f(u) of a stationary time series, the Parzen window (or any other window) spectral density can be regarded as having been superceded by the autoregressive spectral estimator. However, window spectral estimators are needed for comparison (or matching) purposes to be sure that the smoothed spectral density chosen as the estimator



possesses all the significant features present in the raw spectra. Window spectral estimators are also needed to detect strict periodicities in the time series.

One computes also the window spectral distribution function

$$\tilde{F}_M(u) = 2\int_0^u \tilde{f}_M(u')du' , \qquad 0 \leq u \leq 0.5$$

which automatically satisfies $\tilde{F}_M(0.5) = 1$. Note that for $v = 0, \pm 1, \pm 2, \ldots$

$$2\int_0^{0.5} \cos(2\pi uv)\, \tilde{f}_M(u)du = \int_0^1 e^{2\pi iuv}\, \tilde{f}_M(u)du = \hat{\rho}(v)k(\tfrac{v}{M}) .$$

6. Autoregressive analysis. An alternative approach to smoothing the periodogram is to fit, for $m = 1, 2, \ldots$, a smooth function $\hat{f}_m(u)$, $0 \leq u \leq 1$ satisfying

$$\int_0^1 e^{2\pi iuv}\, \hat{f}_m(u)du = \hat{\rho}(v) , \qquad v = 0, \pm 1, \ldots, \pm m ,$$

and among all functions $\hat{f}(u)$ satisfying this constraint

maximizing $\int_0^1 \log \hat{f}(u)$ .

We call $\hat{f}_m(u)$ the autoregressive spectral estimator of order m, since it can be shown that $\hat{f}_m(u)$ may be found by introducing a norm $\tilde{K}_m(g)$ on polynomials g(z) of degree m in a complex variable z whose constant term is constrained to equal 1:

$$g(z) = 1 + a_1 z + \ldots + a_m z^m ,$$

$$\tilde{K}_m(g) = \int_0^1 |g(e^{2\pi iuv})|^2\, \tilde{f}(u)du .$$

Denote by $\hat{g}_m$ the polynomial minimizing $\tilde{K}_m(g)$, and write it explicitly

$$\hat{g}_m(z) = 1 + \hat{a}_{1,m} z + \ldots + \hat{a}_{m,m} z^m ;$$

the coefficients $\hat{a}_{1,m}, \ldots, \hat{a}_{m,m}$ are called the autoregressive coefficients of order $m$, and are computed by solving the Yule Walker equations (for which fast algorithms are available). Let

$$\hat{K}_m = K_m(\hat{g}_m) ;$$

one may show that

$$\hat{K}_m = 1 + \hat{a}_{1,m}\hat{\rho}(1) + \ldots + \hat{a}_{m,m}\hat{\rho}(m).$$

The <u>autoregressive spectral estimator</u> is defined by

$$\hat{f}_m(u) = \hat{K}_m |\hat{g}_m(e^{2\pi i u})|^{-2} .$$

The autoregressive spectral estimator is the spectral density of a time series $Y(t)$ satisfying the autoregressive scheme

$$Y(t) + \hat{a}_{1,m} Y(t-1) + \ldots + \hat{a}_{m,m} Y(t-m) = \varepsilon(t),$$

where $\varepsilon(t)$ is a white noise (zero mean Gaussian) time series with variance $\mathrm{Var}[\varepsilon(t)] = \hat{K}_m$. However, we are using autoregressive schemes not as the true model but as an approximating model.

Associated with the autoregressive spectral estimator $\hat{f}_m(u)$ is a basic problem: choose an order, denoted $\hat{m}$, so that $\hat{f}_{\hat{m}}(u)$ is regarded as "optimal" estimator of $f(u)$.



One approach is to compare the autoregressive spectral distribution function

$$\hat{F}_m(u) = 2 \int_0^u \hat{f}_m(u')du' , \qquad 0 \le u \le 0.5$$

to the raw spectral distribution function $\hat{F}(u)$. One would choose $\hat{m}$ as the smallest value of $m$ for which $\hat{F}_m(u)$ parsimoniously fits $\hat{F}(u)$.

Another approach is to compute and plot various functions of $m$ which describe properties of successive approximating schemes. One can interpret $\hat{a}_{1,m}, \ldots, \hat{a}_{m,m}$ as estimators of the coefficients $a_{1,m}, \ldots, a_{m,m}$ of the minimum mean square error memory $m$ predictor of $Y(t)$ given $Y(t-1), \ldots, Y(t-m)$:

$$-Y^{\mu,m}(t) := -E[Y(t)|Y(t-1), \ldots, Y(t-m)] = a_{1,m}Y(t-1) + \ldots + a_{m,m}Y(t-m) .$$

Next $K_m$ is an estimator of the mean square prediction error

$$\sigma_m^2 = E(|Y(t) - Y^{\mu,m}(t)|^2) ;$$

we call it the biased residual variance (BRV). A more unbiased estimator of $\sigma_m^2$, called the unbiased residual variance (URV), is

$$\hat{\sigma}_m^2 = (1 - \tfrac{m}{T})^{-1} \hat{K}_m .$$

The coefficient $a_{m,m}$ may be shown to be equal to the partial correlation coefficient between $Y(t)$ and $Y(t-m)$, conditioned on $Y(t-1), \ldots, Y(t-(m-1))$. If the time series is in fact an autoregressive scheme of order $m_0$, then $a_{m,m} = 0$ for $m > m_0$. Consequently, an "optimal" order

$\hat{m}$ could be inferred as a value beyond which $\hat{\alpha}_{m,m}$ is not significantly different from zero (say, lies within the limits $\pm 2/\sqrt{T}$). Another criterion for choosing an "optimal" order $\hat{m}$ is the value of m at which the unbiased residual variance $\hat{\sigma}_m^2$ is minimized.

Order determination criteria with stronger theoretical justifications have been introduced by Akaike and Parzen. This program uses the criterion introduced by Parzen (1974), (1977). Choose $\hat{m}$ as the value of m at which CAT (criterion autoregressive transfer) achieves its absolute minimum or its next highest local minimum; CAT is a function of order m defined by, for m = 1, 2, ...,

$$CAT(m) = \frac{1}{T} \sum_{j=1}^{m} \hat{\sigma}_j^{-2} - \hat{\sigma}_m^{-2}$$

The significance level at which the time series is tested to be white noise is achieved by the definition given to CAT(0); we take

$$CAT(0) = -(1 + \frac{1}{T}) .$$

The autoregressive analysis has an extensive output. It plots partial autocorrelations, biased residual variance, unbiased residual variance, and CAT. It prints out a best order and second best order for approximating autoregressive schemes. It prints out their coefficients, their spectral densities, and their spectral distribution functions.

7. Select Autoregressive analysis. One may next determine autoregressive schemes of the form

$$Y(t) + \alpha_{j_1} Y(t - j_1) + \ldots + \alpha_{j_k} Y(t - j_k) = \varepsilon(t)$$

in which occur only lags $j_1, \ldots, j_k$ whose coefficients $\alpha_{j_1}, \ldots, \alpha_{j_k}$ are "significantly" different from zero. Such



schemes help us interpret as a time series decomposition the whitening filter found in the autoregressive analysis. Granger and Newbold (1977), p. 176, give inadequate references to the history of stepwise autoregression. It was extensively used in Parzen (1964). The earliest reference may be MacDonald and Ward (1963).

8. Plots of predictor time series. As calculated by Newton and Pagano (1979).

9. Seasonally (or periodically) stationary time series models. Described in Parzen and Pagano (1979).

## 6. Non-stationary Analysis

There are many ways in which a time series can be non-stationary; consequently there are several data analytic approaches to deciding whether non-stationarity is present and how to model it.

1. Time series plots and regression model. The time series plot may clearly indicate a trend (represented by a linear or polynomial function of time) and periodicities (represented by linear combinations of cosine and sine functions of suitable frequencies). One might then adopt a regression model.

2. Correlations. A condition for a stationary time series to be ergodic is $\Sigma|\rho(v)| < \infty$; this condition cannot be checked, but the manner in which the sample correlations $\hat{\rho}(v)$ are decaying to 0 as v increases might distinguish between stationary and non-stationary time series. On the basis of the plot of $\hat{\rho}(v)$ one can classify a sample correlation in three types:

(i) very slow decay to 0: non-stationary,

(ii) very fast decay to 0: whitened or contracted,

(iii) neither of the above: stationary but not white noise.

These three types are illustrated in Figure A (using the famous international airline passenger time series).

3. Prediction residual variance. A stationary time series with a spectral density function f(u) is called non-predictable if $\sigma_\infty^2 > 0$ (where $\sigma_\infty^2$ is the infinite memory one-step ahead mean square prediction error), or equivalently if $\int_0^1 \log f(u)du$ is finite. Otherwise it is called predictable. I recommend using in practice the following equivalences:

$$\text{non-predictable} \equiv \text{stationary}$$

$$\text{predictable} \equiv \text{non-stationary}$$

To determine in practice whether a time series is non-predictable or predictable, one examines the successive autoregressive memory m mean square prediction errors $\hat{\sigma}_m^2$. We consider an observed time series to be predictable (non-stationary) if (i) for some value of m

$$\sigma_m^2 < \frac{8}{T},$$

or (ii) more generally there exists a "naive prediction error"

$$\tilde{Y}(t) = Y(t) + a_1 Y(t-1) + \ldots + a_m Y(t-m)$$

satisfying

$$\frac{\Sigma|\tilde{Y}(t)|^2}{\Sigma|Y(t)|^2} < \frac{8}{T}.$$



We call $\tilde{Y}(t)$ a "naive predictor error" because it could arise from a naive predictor as follows. If a time series has two obvious periods $\lambda_1$ and $\lambda_2$, one might predict Y(t) by

$$Y^{\text{naive}}(t) = Y(t-\lambda_1) + Y(t-\lambda_2) - Y(t-\lambda_1-\lambda_2).$$

For monthly data with trend and seasonal one would use $\lambda_1 = 1$, $\lambda_2 = 12$, and

$$Y^{\text{naive}}(t) = Y(t-1) + Y(t-12) - Y(t-13).$$

The naive prediction error is

$$\begin{aligned}\tilde{Y}(t) &= Y(t) - Y^{\text{naive}}(t)\\ &= Y(t) - Y(t-\lambda_1) - Y(t-\lambda_2) + Y(t-\lambda_1-\lambda_2)\\ &= (I - L^{\lambda_1})(I - L^{\lambda_2})Y(t)\\ &= \nabla_{\lambda_1}\nabla_{\lambda_2}Y(t)\end{aligned}$$

where L is the backward lag operator (defined by $L^\lambda Y(t) = Y(t-\lambda)$ for any integer $\lambda$) and $\nabla_\lambda = I - L^\lambda$ is the $\lambda$-th difference operator.

The notion of the "naive prediction error" provides an interpretation of a method proposed, and made popular, by Box and Jenkins (1970) for transforming a non-stationary time series to a stationary time series. They propose first differencing d times and $\lambda$-th differencing D times to form

$$(I - L)^d (I - L^\lambda)^D Y(t).$$

They conceive that d and D could be any one of the integers 1, 2, ... although in practice one would rarely choose values greater than 2.

To determine values of d and D, one examines the sample correlations of various differenced time series and choose one which is almost whitened.

My use of differencing as part of a time series modeling strategy is quite different. The need for differencing is indicated by the stationary analysis if the residual variance $\hat{\sigma}_a^2$ of best fitting schemes is of the order of 8/T. How to difference is indicated by the location of spectral peaks in spectral estimators which are to be annihilated by the type of differencing chosen. One seeks to use the gentlest degree of differencing in order to preserve the spectral structure as much as possible to enable our automatic stationary modeling program maximum opportunity to choose a model that describes all the features of the data.

Instead of applying $I - L^{12}$ (twelth differencing) to monthly data one should apply in succession its factors (the frequencies they "annihilate" are listed at right)

| | |
|---|---|
| $I - L$ | $u = 0$ |
| $I - \sqrt{3}\,L + L^2$, | $u = \frac{1}{12}$ |
| $I - L + L^2$, | $u = \frac{2}{12}$ |
| $I + L^2$, | $u = \frac{3}{12}$ |
| $I + L + L^2$, | $u = \frac{4}{12}$ |
| $I + \sqrt{3}\,L + L^2$, | $u = \frac{5}{12}$ |
| $I + L$ | $u = \frac{6}{12}$ |



If spectral peaks are present at all the harmonics of $\frac{2\pi}{12}$, one would apply $I - L^{12}$. But it seems hard to imagine when one would want to apply both $I - L$ and $I - L^{12}$.

Monthly data whose cycle is quarterly (period three months) might be transformed to stationarity by applying $I + L + L^2$ or

$$(I - L^3) = (I - L)(I + L + L^2)$$

Quarterly data whose cycle is annual have period 4; they might be transformed to stationarity by applying $I - L^4$ or its factors $I - L$, $I + L^2$, $I + L$.

7. International Airline Data

A monthly time series of monthly passenger totals in international air travel from 1949 to 1960 (which has length T = 144) has been considered by many authors as an illustrative example of time series since it is presented as one of the test series in Box and Jenkins (1970). The model fitted by the latter,

$$\nabla_{12}\nabla_1 \log Y(t) = (I - \theta_1 L)(I - \theta_{12} L^{12})\varepsilon(t) ,$$

is called "the airline model." For the airline data, Box and Jenkins estimate $\theta_1 = 0.4$, $\theta_{12} = 0.6$, $\sigma_\varepsilon^2 = 0.0013$. A model fitted by Parzen (1979), but which is in fact compatible with the airline model, is

$$g_{13}(L)\nabla_{12} \log Y(t) = \varepsilon(t)$$

where $\sigma_\varepsilon^2 = 0.00127$, and $g_{13}(L)$ is a lag polynomial of degree 13, which can be written approximately

$$g_{13}(L) = 1 - 0.74L + 0.38L^{12} - 0.31L^{13} \doteq (I - .74L)(I + .38L^{12})$$

The airline data is also discussed by Makradakis (1978) and Chatfield (1978). Chatfield considered forecasts based on fitting an AR(12) to $\nabla_1 Y(t)$ whereas the best fitting AR scheme using CAT has order 14.

Parzen derives his model using an iteration as follows. (I) Autoregressive analysis of $Z(t) = \log Y(t)$ indicates predictability, and $\hat{m} = 13$. The coefficients suggest forming $\nabla Z(t)$. (II) However, $\tilde{Z}(t) = \nabla_{12} Z(t)$ is chosen as the time series to be analyzed at the next iteration. Its autoregressive analysis indicates that it is unpredictable. Applying $\nabla_{12}$ rather than $\nabla$ seems wiser in general as the lowest frequency spectrum is less affected. There seems no reason to apply $\nabla_1 \nabla_{12}$ as either operator alone suffices to transform $Z(t)$ to an unpredictable (stationary) time series. (III) The last iteration is a stepwise autoregression of $\tilde{Z}(t)$.

The three types of time series (predictable or non-stationary, unpredictable or stationary, whitened) could be inferred from the decay of the sample correlograms or from the range of values of the CAT order-determined autoregressive spectral density. These graphs for the various differenced versions of $Z(t)$ are given in Figure A.

## 8. Metals Series of Makridakis

Makridakis (1978) presents an interesting monthly series for analysis, which he calls the metal series; it consists of 144 monthly values of carbon steel monthly shipments from 1961-1972. Some researchers in time series analysis apply the "airline model" (defined in the previous section) to any monthly economic time series. Makradakis fits the metal series by the model

$$\nabla Y(t) = (I - \theta_1 L)(I - \theta_{12} L^{12})\varepsilon(t)$$



where $\theta_1 = .45$, $\theta_2 = -.26$. He reports as a paradox that this presumably "optimum" model does not forecast as well as the random walk model $\nabla Y(t) = \varepsilon(t)$. I believe the resolution of this paradox is to find a more appropriate model. An AR(2) for $\tilde{Y}(t) = Y(t) - 3082$ is

$$\tilde{Y}(t) - .44\tilde{Y}(t-1) - .35\tilde{Y}(t-2) = \varepsilon(t) .$$

CAT selected AR(2) as best fitting and AR(13) as second best fitting. The two models have quite different spectral densities and they do not forecast similarly. The spectral distribution functions comparison leads one to prefer the AR(13) model, which was

$$\begin{aligned}\tilde{Y}(t) - .46\tilde{Y}(t-1) - .30\tilde{Y}(t-2) - .05\tilde{Y}(t-3) + .10\tilde{Y}(t-5) \\ + .02\tilde{Y}(t-6) - .08\tilde{Y}(t-7) - .04\tilde{Y}(t-8) + .01\tilde{Y}(t-9) \\ + .01\tilde{Y}(t-10) + .02\tilde{Y}(t-11) - .29\tilde{Y}(t-12) + .21\tilde{Y}(t-13) = \varepsilon(t)\end{aligned}$$

The models obtained by applying the stationary analysis (described in Section 4) to the original time series, to its logaritha, and to its square root were essentially the same. The results of various stages of the analysis are presented in Table I and Figure B.

Many of the coefficients in the above AR(13) model are not significantly different from zero which suggests that an AR model be determined by select or subset autoregression. The "best" select AR model is the AR(2) model. Based on spectral distributions we prefer an AR(13) and therefore fit a "best" AR scheme with 4 coefficients; it was

$$\tilde{Y}(t) - .47\tilde{Y}(t-1) - .30\tilde{Y}(t-2) - .30\tilde{Y}(t-12) + .22\tilde{Y}(t-13) = \varepsilon(t) ,$$

with residual variance (MSE) 0.44. We denote this model AR(1, 2, 12, 13).

To demonstrate that a better fitting model is obtained if one does not first difference, let us compare mean average percentage forecast errors using AR models with the results reported by Makridakis (1978), p. 273.

Mean Average Percentage Error (MAPE)

for Various Forecasting Horizons

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| AR(1, 2, 12, 13) for $\tilde{Y} = Y - \bar{Y}$ | 8.65 | 9.86 | 11.04 | 11.76 | 12.20 | 12.31 |
| AR(2) for $\tilde{Y} = Y - \bar{Y}$ | 9.75 | 10.68 | 11.82 | 12.46 | 12.59 | 12.59 |
| AR(2) for $\tilde{Y} = Y -$ monthly mean | 7.63 | 8.03 | 9.21 | 10.09 | 10.54 | 10.63 |
| $\nabla Y = \epsilon$ | 10.88 | 13.55 | 15.87 | 18.05 | 18.77 | 18.84 |
| $\nabla Y = (I - \theta_1 L)(I - \theta_{12} L)\epsilon$, $\theta_1 = .45$, $\theta_2 = -.26$ | 11.42 | 13.21 | 14.87 | 15.22 | 15.66 | 16.31 |
| | 7 | 8 | 9 | 10 | 11 | 12 |
| AR(1, 2, 12, 13) for $\tilde{Y} = Y - \bar{Y}$ | 12.45 | 12.48 | 12.40 | 12.30 | 12.17 | 12.04 |
| AR(2) for $\tilde{Y} = Y - \bar{Y}$ | 12.50 | 12.60 | 12.67 | 12.61 | 12.57 | 12.62 |
| AR(2) for $\tilde{Y} = Y -$ monthly mean | 11.00 | 11.41 | 11.51 | 11.70 | 11.83 | 11.82 |
| $\nabla Y = \epsilon$ | | | 18.60 | | | 18.45 |
| $\nabla Y = (I - \theta_1 L)(I - \theta_{12} L)$ | | | 17.03 | | | 18.49 |

The metals series given by Makridakis (1978) seems to be a good series for forecasting experts to use for a competitive comparison of various forecasting methods. If we are to achieve a general approach to statistical forecasting, we must deal with the following fact: parameters of time series models that do not appear statistically significant when model fitting using conventional statistical tests usually do not have a statistically



significant effect on the forecast errors of the model, but such parameters may have a scientifically significant effect on the spectral density of the time series, which could be used to guide our model building strategy.

Another approach to modeling the metals series is to estimate the monthly means, and apply a stationary analysis to the time series $\tilde{Y}(t)$ of residuals from the monthly means. The monthly means do not test out to be significantly different from each other; nevertheless, the forecasting errors obtained by modeling the time series as "Y(t) - monthly means is an AR(2)" seem to have smaller MAPE than those from the model "Y(t) - mean is an AR(1, 2, 12, 13)." This conclusion is not be accepted as conclusive because our work used the same data base to fit monthly means as was used to generate forecast errors. One conjectures that the AR(1, 2, 12, 13) model provides more robust forecasts.

Our time series analysis computes mean square forecast error (MSE), denoted $\sigma^2$, expressed as a proportion of the original mean square of the time series. It is of interest to relate MAPE and MSE.

The metals data has mean 3082 and standard deviation 571. The MSE, or unbiased residual variance $\hat{\sigma}^2$ of AR(2) is .49 (measured as a proportion of the variance of the original time series); its square root $\hat{\sigma}$ (equal to .7) is called the unbiased residual standard deviation. It has the interpretation that the forecast errors have standard deviation approximately equal to (.7)(571) = 400. The coefficient of variation 400/3082 = 13% could be regarded as providing an upper bound to MAPE (whose actual value is 9.75%).

The relation in general can be expressed MAPE $\leq$ MAPE BOUND, defining

MAPE BOUND = (Coefficient of variation of original data) × $\sqrt{MSE}$

The bound is <u>not</u> a theorem; it seems to apply to non-negative data with a small coefficient of variation.

9. How well can stationary time series models predict?

Plots of time series and their one-step ahead predictors (such as the airline series in Figure A) often seem to indicate impressive evidence of time series analysis' ability to predict. However, this impression is correct only for predictable time series (which in our terminology are non-stationary time series). For stationary or non-predictable time series one expects the predicted series to have the same mean as the original series and to have standard deviation equal to a proportion $\hat{\sigma}$ of the original standard deviation, where $\hat{\sigma}$ is the residual standard deviation. Predicted values rarely attain the extreme values attained by the original time series and therefore appear to seriously underpredict large deviations from the mean. This property seems to be to be inherent in stationary time series modeling. If one desires a model that is able to predict extreme values of the time series, one will have to use other kinds of models (such as point process models for the times at which extreme values occur).

The fact that the predicted values of a stationary time series have a more concentrated distribution than the original values is illustrated by comparing the histograms of these two data sets (Table II) for the metals series.



Table I

Autoregressive Analysis of Metals Series

Sample Size T = 144, Predictability Threshold $\hat{\sigma}/T$ = .055

| | Original Time Series | Square Root | Logarithm |
|---|---|---|---|
| Mean | 3082 | 55.3 | 8.02 |
| Standard deviation | 569 | 5.1 | .18 |
| Coefficient of variation | .18 | .09 | .02 |
| Best AR order $\hat{m}$ | 2 | 2 | 2 |
| CAT($\hat{m}$) | -2.01 | -2.02 | -2.03 |
| MSE $\hat{\sigma}^2$ | .49 | .49 | .49 |
| MAPE Bound | .12 | .06 | .014 |
| AR Coefficients | 1 -.44 | 1 -.45 | 1 -.47 |
| | 2 -.35 | 2 -.33 | 2 -.32 |
| Second Best AR order $\hat{m}$ | 13 | 13 | 13 |
| CAT($\hat{m}$) | -1.93 | -1.94 | -1.94 |
| MSE $\hat{\sigma}^2$ | .47 | .47 | .47 |
| MAPE Bound | .12 | .06 | .014 |
| AR Coefficients | 1 -.46 | -.48 | -.50 |
| | 2 -.30 | -.30 | -.30 |
| | 3 -.05 | -.02 | .01 |
| | 4 -.00 | .00 | .01 |
| | 5 .10 | .08 | .06 |
| | 6 .02 | .03 | .04 |
| | 7 -.08 | -.09 | -.10 |
| | 8 -.04 | -.04 | -.03 |
| | 9 .01 | .02 | .02 |
| | 10 .01 | .01 | .01 |
| | 11 .02 | .02 | .02 |
| | 12 -.29 | -.30 | -.29 |
| | 13 .21 | .20 | .19 |

Table II

Histogram of Metals Time Series and One-Step Predictions

| | Original | Predicted |
|---|---|---|
| 4551-4650 | 2 | |
| 4451-4550 | 1 | |
| 4351-4450 | - | |
| 4251-4350 | 1 | |
| 4151-4250 | 2 | 1 |
| 4051-4150 | 3 | 2 |
| 3951-4050 | 2 | - |
| 3851-3950 | 5 | 2 |
| 3751-3850 | 3 | 5 |
| 3651-3750 | 4 | 3 |
| 3551-3650 | 4 | 7 |
| 3451-3550 | 7 | 7 |
| 3351-3450 | 7 | 12 |
| 3251-3350 | 11 | 13 |
| 3151-3250 | 9 | 9 |
| 3051-3150 | 8 | 8 |
| 2951-3050 | 12 | 16 |
| 2851-2950 | 8 | 14 |
| 2751-2850 | 7 | 7 |
| 2651-2750 | 9 | 16 |
| 2551-2651 | 9 | 9 |
| 2451-2550 | 7 | 6 |
| 2351-2450 | 8 | 1 |
| 2251-2350 | 8 | 2 |
| 2151-2250 | 1 | 2 |
| 2051-2150 | 1 | |
| 1951-2050 | 2 | |
| 1851-1950 | 1 | |



If a time series is predictable, we interpret it to be because of the presence of a trend or cyclic components. If a time series is non-predictable, it is because "regularities" observed in the past have not recurred in the present, and innovations (random shocks) dominate the behavior of the time series; we call such a time series a stationary time series. It is non-predictable in the sense that partial prediction is possible only over a limited future, and the predicted values will always underpredict extreme values of the original time series.

## 10. Diagnostic Checks for Time Series Models

One of the important problems of time series modeling is assessing the adequacy of a proposed model. Since a time series model is equivalent to a transformation of the time series to white noise, a basic approach to goodness of fit of a model is to generate the time series of "residuals" and test whether they are distributed as white noise. One can base such tests on: (1) the sample correlations of the residuals; (2) the sample spectral distribution function of the residuals; and (3) the best fitting autoregressive scheme for the residuals. In my judgement, the last test is the most powerful.

A currently popular test is the so-called "portmanteau" test, which treats the sum of squares of the first m sample correlations of the residuals as chi-squared distributed with m degrees of freedom under the null hypothesis of white noise. Often one chooses m = 20 [see, for example, Granger and Newbold (1977), p. 106]. This test seems to almost always accept the hypothesis of white noise. To test whether 20 correlations are not significantly different from zero one needs a "subset regression" or

"model identification" type statistical procedure which considers as null hypotheses all possible hypotheses of the form: a specified subset of the 20 parameters equals 0.

The inadequacy of diagnostic tests for models which are based solely on testing residuals for whiteness is pointed out by Kashyap and Rao (1976), pp. 283-4. "Consider the best fitting AR(2) [to the lynx and sunspot series]. The residuals from them pass all the whiteness tests, but the correlograms of the outputs of these models are bad fits to the corresponding empirical correlograms. Hence we have to try other classes until the correct class is identified." Our approach recommends matching spectral distribution functions rather than correlograms.

## 11. ARMA Models

Autoregressive modeling as described in this paper provides a technique for two basic problems of time series modeling: (1) for a general time series, determining whether it is non-stationary (predictable) or stationary (non-predictable); and (2) for a stationary time series, determining the whitening filter which transforms it to its innovations (white noise).

When modeling a stationary time series Y(t), it is desirable to also find the best fitting autoregressive - moving average (ARMA) representation

$$Y(t) + \alpha_1 Y(t-1) + \ldots \alpha_p Y(t-p) = \varepsilon(t) + \beta_1 \varepsilon(t-1)$$

$$+ \ldots + \beta_q \varepsilon(t-q) \ .$$

An interesting technique for finding the orders p and q has been introduced by Gray (1978).



ARMA models are widely used by followers of the methods of Box and Jenkins (1970). It seems to me that many of these applications are motivated not by the interpretability of the MA part of the model, but by the fact that an ARMA filter with a few parameters may provide a parsimonious parametrization of the infinite autoregressive filter useful for estimating its transfer function. However, the latter can be estimated directly by using AR schmes whose order is determined adaptively from the data rather than by guessing an ARMA model to fit the data.

## 12. Acknowledgement

The theory described in this paper could not have been developed without the extensive help of my colleague Professor H. J. Newton of Texas A&M University and the library TIMESBOARD of time series computing programs which he has implemented.

## References


Anderson, O. D. (1977). "A commentary on a 'A Survey of Time Series,'" International Statistical Review, 45, 273-297.

Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control. Holden Day: San Francisco.

Chatfield, C. (1978). "Adaptive filtering: A critical assessment," J. Opr. Res. Soc., 29, 891-896.

Granger, C. W. J. and Newbold, P. J. (1977). Forecasting Economics Time Series. Academic Press: New York.

Gray, H. L., Kelley, G. D., and McIntire, D. P. (1978). "A new approach to ARMA modeling," Communications Statistics, Simulation and Computation B7, 1-77.

Kashyap, R. L. and Rao, A. R. (1976). Dynamic Stochastic Models for Empirical Data, Academic Press: New York.

McDonald, H. J. and Ward, F. (1963). "The prediction of geomagnetic disturbance indices," Journal of Geophysical Research, 68, 3351-3373.

Makridakis, S. (1978). "Time series analysis and forecasting: An update and evaluation," International Statistical Review, 46, 255-278.

Makridakis, S. and Hibon, M. (1979). "Accuracy of forecasting: An empirical investigation," J. R. Statist. Soc. A, 142, Part 2.

Newton, H. J. and Pagano, M. (1979). "Least squares prediction for mixed autoregressive moving average time series," Technical Report N-2, Institute of Statistics, Texas A&M University.

Pagano, M. (1979). "Some recent advances in autoregressive processes," Time Series Analysis, Proceedings of Meeting at Iowa, sponsored by Institute of Mathematical Statistics.





Parzen, E. (1964). "An approach to empirical time series analysis," J. Res. Nat. Bur. Standards, 68D, 937-951.

Parzen, E. (1967). Time Series Analysis Papers, Holden-Day: San Francisco.

Parzen, E. (1967). "Empirical multiple time series analysis," Fifth Berkeley Symposium, University of California Press, I, 305-340.

Parzen, E. (1967). "The role of spectral analysis in time series analysis," Review of the International Statistical Institute, 35, 125-141.

Parzen, E. (1969). "Multiple time series modeling," Multivariate Analysis - II, edited by P. Krishnaiah, Academic Press: New York, 389-409.

Parzen, E. (1969). "Statistical spectral analysis (single channel case) in 1968," Proceedings of NATO Advanced Study Institute on Signal Processing, Vol. 1, 14. 1-14. 20.

Parzen, E. and Pagano, M. (1973). "Timesboard: A time series package," Proceedings of Computer Science and Statistics 7th Annual Symposium on the Interface, Iowa State University, 193-197.

Parzen, E. (1974). "Some recent advances in time series modeling," IEEE Transactions on Automatic Control, AC-19, 723-730.

Parzen, E. (1975). "Some solutions of the time series modeling and prediction problem," The Search for Oil: Proceedings of a Symposium on Time Series Analysis, edited by D. Owen. Marcel Dekker: New York, 1-16.

Parzen, E. (1976). "An approach to time series modeling and forecasting illustrated by hourly electricity demands," Technical Report.

Parzen, E. (1977). "Multiple time series: Determining the order of approximating autoregressive schemes," Multivariate Analysis - IV, edited by P. Krishnaiah, North Holland: Amsterdam, 283-295.

Parzen, E. and Pagano, M. (1979). "An approach to modeling seasonally stationary time series," Journal of Econometrics, 9, 137-153.

Parzen, E. and Newton, H. J. (1979). "Multiple time series modeling, II," Multivariate Analysis - V, edited by P. Krishnaiah.

Parzen, E. (1979). "Forecasting and whitening filter estimation," Management Science, Special Issue on Forecasting, ed. Makridakis, S. and Wheelwright, S.

Parzen, E. (1979). "Nonparametric Statistical Data Modeling," Journal of the American Statistical Association, 74.

Yule, G. U. (1921). "On the time-correlation problem, with special reference to the variate-difference correlation method. J. Roy. Stat. Soc., 84, 497-537.



Figure A

Let $Z(t)$ denote the time series of logarithms of international airline passengers, $\nabla Z(t)$ its 1 difference, $\nabla_{12} Z(t) = Z(t) - Z(t-12)$ its 12 difference, and $\nabla \nabla_{12} Z(t)$ its 1, 12 difference.

Correlations indicate three types; slow decay indicates non-stationary or predictable ($Z$ and $\nabla Z$), quick decay to 0 could indicate white noise but in general is taken as an indicator of stationary ($\nabla \nabla_{12} Z(t)$ is in our judgement whitened too much, $\nabla_{12} Z(t)$ is stationary).

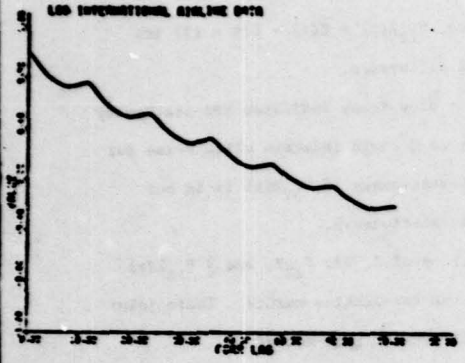
Partial correlations and correlations of $Z$, $\nabla Z$, $\nabla_{12} Z$, and $\nabla \nabla_{12} Z(t)$ are the model identification tools of the Box-Jenkins method. Their joint interpretation is best left to practicioners of that method.

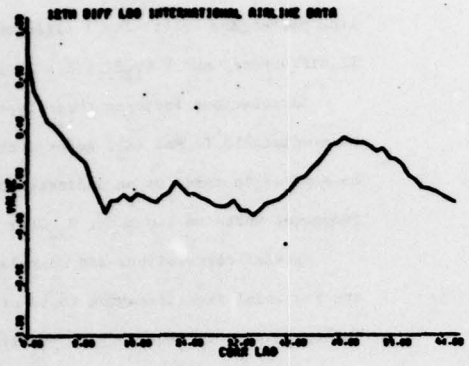
CAT is an autoregressive order determining tool; the orders at which it has an abeolute minimum and the next lowest relative minimum are regarded as best orders for investigation as to how well the autoregressive spectral distribution functions of these orders match the raw spectral distribution function (cumulative periodogram). A CAT function with a sharp dip indicates non-stationarity.

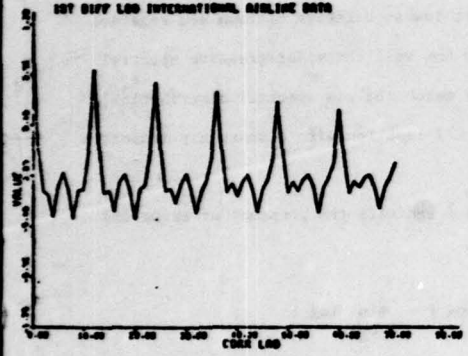
Autoregressive spectral densities $\hat{f}$ indicate the presence of trend and seasonal periods by a large value of

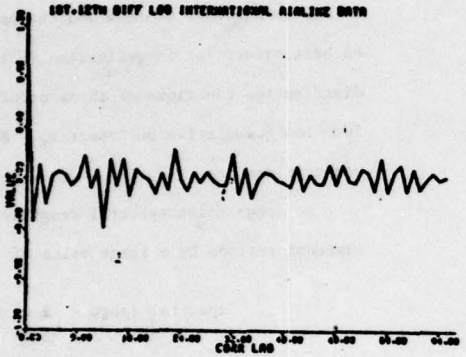
$$\text{spectral range} = \max_{0 \leq u \leq 1} \log \hat{f} - \min_{0 \leq u \leq 1} \log \hat{f}$$

and by the cusp-like character of $\hat{f}$ at its relative maxima (or bumps). White noise is indicated by $\hat{f}$ almost constant (spectral range $\doteq 0$). Stationarity but not white is indicated by moderate values of spectral range.

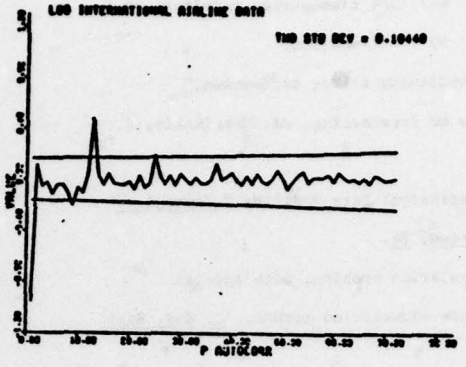

LOG INTERNATIONAL AIRLINE DATA
VALUE
CORR LAG

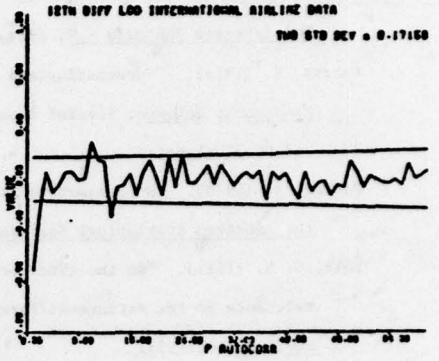

12TH DIFF LOG INTERNATIONAL AIRLINE DATA
VALUE
CORR LAG

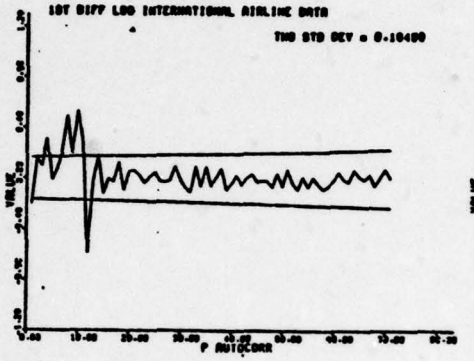

1ST DIFF LOG INTERNATIONAL AIRLINE DATA
VALUE
CORR LAG

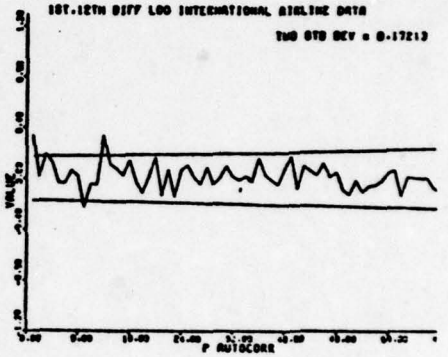

1ST.12TH DIFF LOG INTERNATIONAL AIRLINE DATA
VALUE
CORR LAG

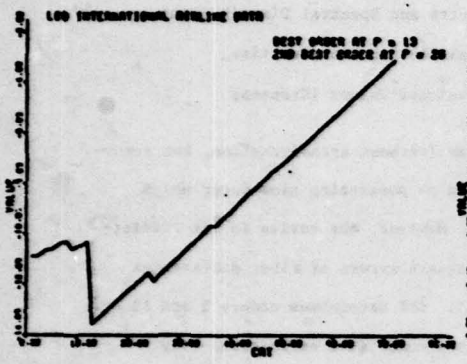


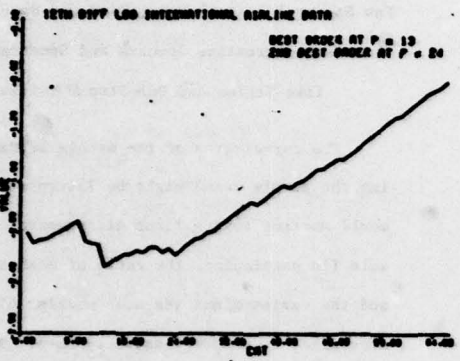

LOG INTERNATIONAL AIRLINE DATA
THO STD DEV = 0.10440
VALUE
P AUTOCORR

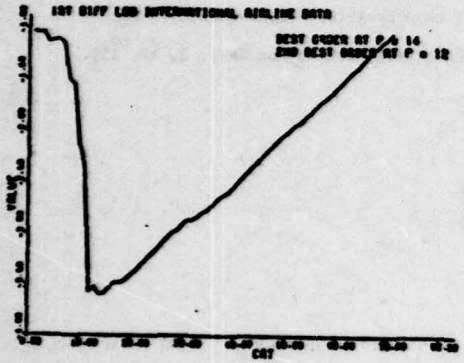

12TH DIFF LOG INTERNATIONAL AIRLINE DATA
THO STD DEV = 0.17160
VALUE
P AUTOCORR

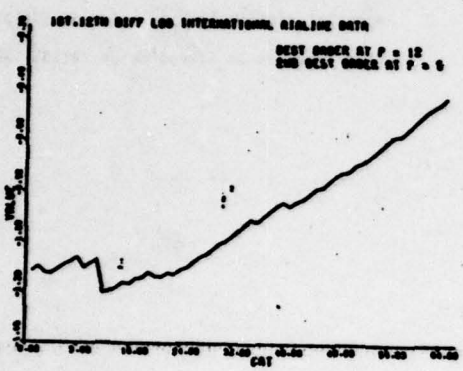

1ST DIFF LOG INTERNATIONAL AIRLINE DATA
THO STD DEV = 0.10490
VALUE
P AUTOCORR


1ST.12TH DIFF LOG INTERNATIONAL AIRLINE DATA
THO STD DEV = 0.17213
VALUE
P AUTOCORR

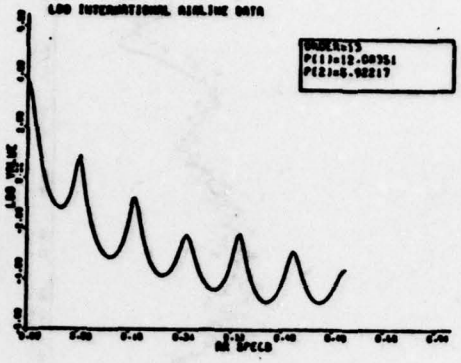

LOG INTERNATIONAL AIRLINE DATA
BEST ORDER AT P = 13
2ND BEST ORDER AT P = 20
LOG VALUE
CAT

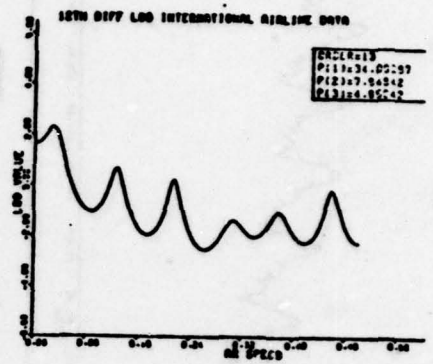

12TH DIFF LOG INTERNATIONAL AIRLINE DATA
BEST ORDER AT P = 13
2ND BEST ORDER AT P = 24
LOG VALUE
CAT

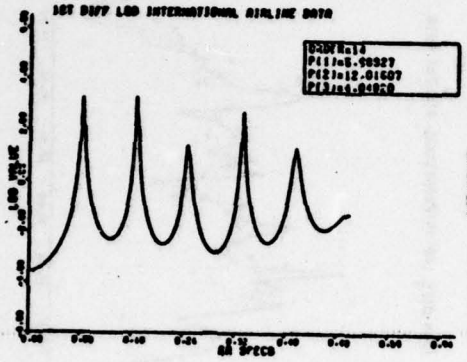

1ST DIFF LOG INTERNATIONAL AIRLINE DATA
BEST ORDER AT P = 14
2ND BEST ORDER AT P = 12
LOG VALUE
CAT

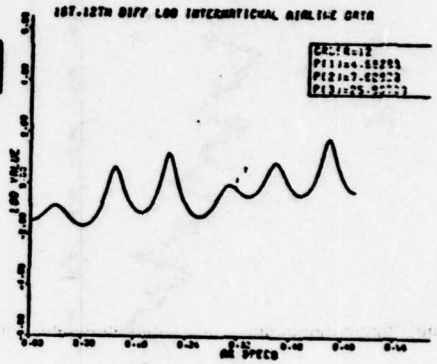

1ST.12TH DIFF LOG INTERNATIONAL AIRLINE DATA
BEST ORDER AT P = 12
2ND BEST ORDER AT P = 0
LOG VALUE
CAT




LOG INTERNATIONAL AIRLINE DATA
ORDER=15
P(1)=12.08351
P(2)=6.98217
LOG VALUE
AR SPECD


12TH DIFF LOG INTERNATIONAL AIRLINE DATA
LOG VALUE
AR SPECD


1ST DIFF LOG INTERNATIONAL AIRLINE DATA
ORDER=14
P(1)=8.08827
P(2)=12.81607
LOG VALUE
AR SPECD


1ST.12TH DIFF LOG INTERNATIONAL AIRLINE DATA
LOG VALUE
AR SPECD

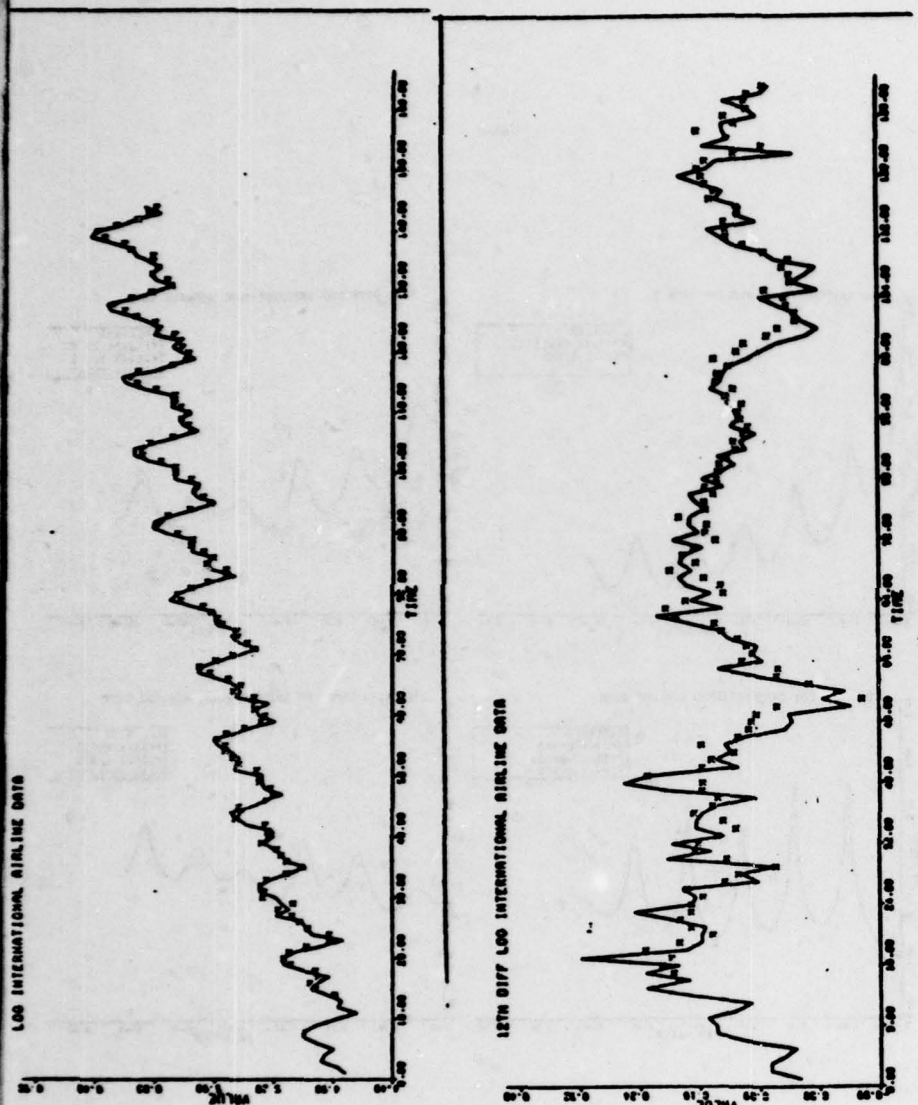


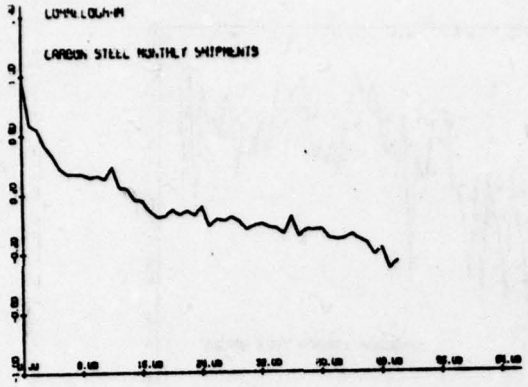


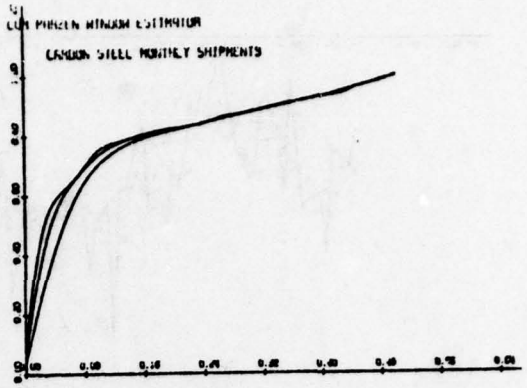


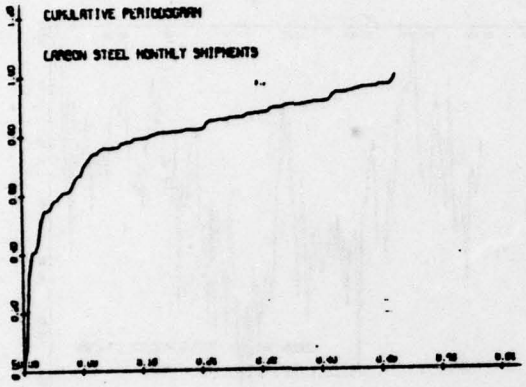

**Figure B**

**Metals Series of Makridakis; Plots of Correlations; Raw Spectral Distribution; Windowed Spectra and Spectral Distribution; Autoregressive Spectra and Spectral Distribution Function; Time Series and One-Step Ahead Predicted Values (Crosses)**

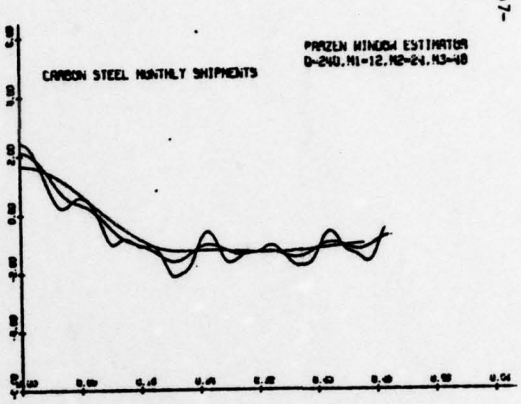
The correlogram of the metals series (without transformation, but removing the sample mean) might be interpreted as possessing slow decay which would warrant taking first differences. However, the series is not predictable (in particular, the ratio of mean square errors of first differences and the series minus its mean equals .6). CAT determines orders 2 and 13 for approximating autoregressive schemes. The cumulative periodogram (raw spectral distribution function) is matched better at low frequencies by the spectral distribution function of AR(13) than by AR(2). Consequently, the metals series is fitted as an AR(13), and equally well by an AR(1, 2, 12, 13).

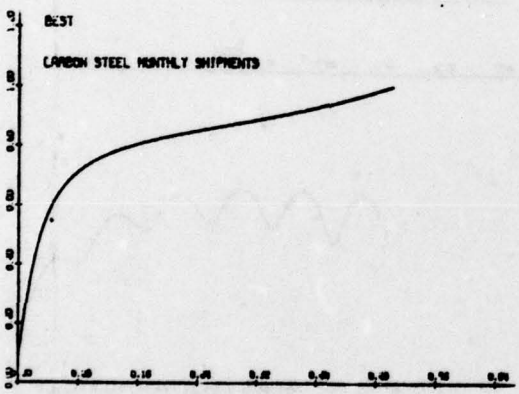

CARBON STEEL MONTHLY SHIPMENTS

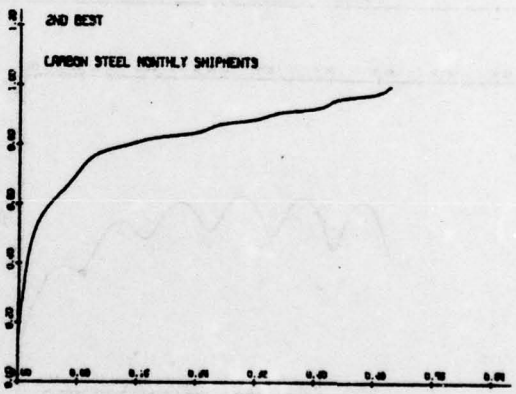

CARBON STEEL MONTHLY SHIPMENTS

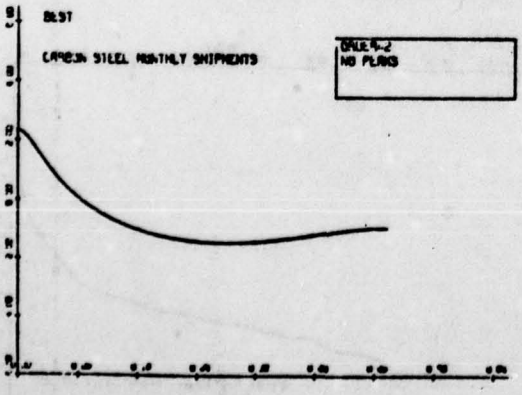

CUMULATIVE PERIODOGRAM
CARBON STEEL MONTHLY SHIPMENTS

-47-

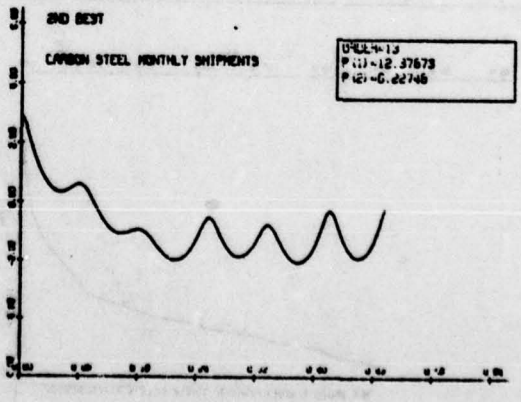

PARZEN WINDOW ESTIMATOR
D=240.M1=12.M2=24.M3=48
CARBON STEEL MONTHLY SHIPMENTS

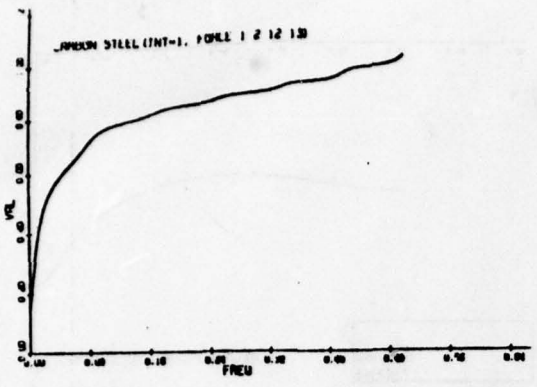

BEST
CARBON STEEL MONTHLY SHIPMENTS

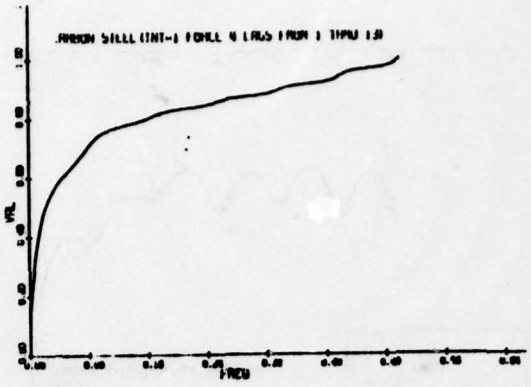

2ND BEST
CARBON STEEL MONTHLY SHIPMENTS


BEST
CARBON STEEL MONTHLY SHIPMENTS
NO PLANS

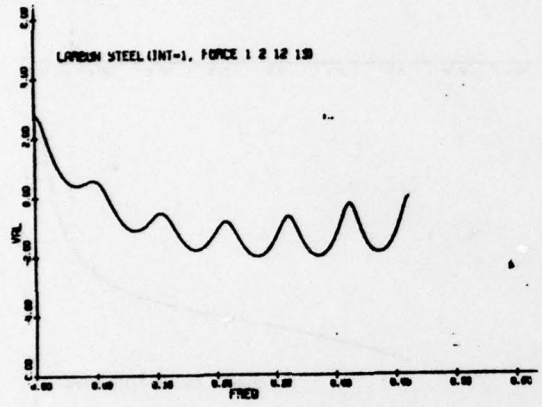
-48-

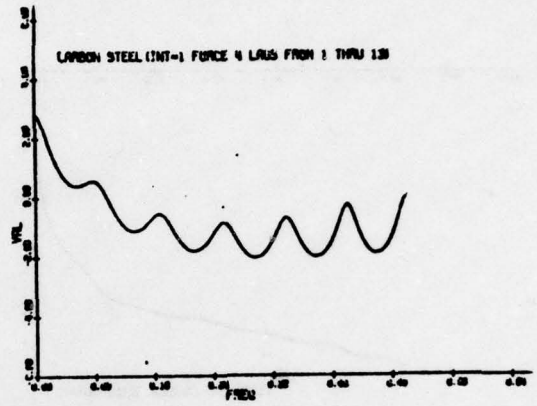

2ND BEST
CARBON STEEL MONTHLY SHIPMENTS

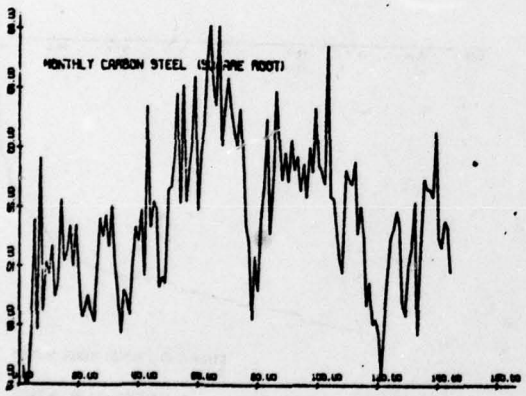

CARBON STEEL (INT=1, FORCE 1 2 12 13
VAL
FREQ


CARBON STEEL (INT=1 FORCE 4 LAGS FROM 1 THRU 13
VAL
FREQ

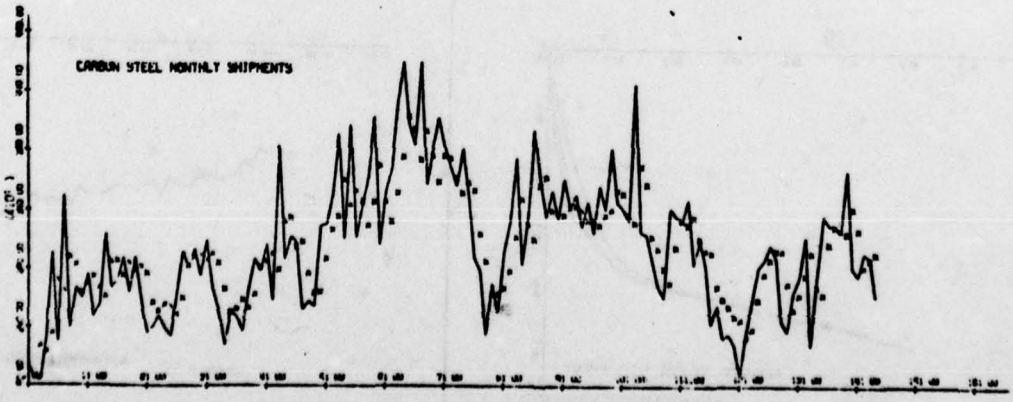
-49-


CARBON STEEL (INT=1, FORCE 1 2 12 13
VAL
FREQ


CARBON STEEL (INT=1 FORCE 4 LAGS FROM 1 THRU 13
VAL
FREQ


MONTHLY CARBON STEEL (SQUARE ROOT)

-50-


CARBON STEEL MONTHLY SHIPMENTS